

Brazilian Logistics: Total Daily Orders

Dataset Background & Structure

This dataset is based on a real database of a Brazilian company of large logistics and collected for 60 days. It includes twelve predictive attributes and the total daily order, which is the target of prediction. The dataset was downloaded from [UCI Machine Learning Repository](#) as a CSV(‘;’ separated), and it contains 13 columns and 60 rows.

Attribute information

@attribute Week_of_the_month {1.0, 2.0, 3.0, 4.0, 5.0}
@attribute Day_of_the_week_(Monday_to_Friday) {2.0, 3.0, 4.0, 5.0, 6.0}
@attribute Non_urgent_order integer
@attribute Urgent_order integer
@attribute Order_type_A integer
@attribute Order_type_B integer
@attribute Order_type_C integer
@attribute Fiscal_sector_orders integer
@attribute Orders_from_the_traffic_controller_sector integer
@attribute Banking_orders_(1) integer
@attribute Banking_orders_(2) integer
@attribute Banking_orders_(3) integer
@attribute Target_(Total_orders) integer

Exploratory Data Analysis Goals

The dataset is relatively small with 60 rows. Still, it is hard to understand the relationship between collected data and distinguish important features for predicting total daily order.

First, before analysis, we will clean data; we will unify data type and replace numberings with category names for better representation during analysis.

Then, we will perform EDA (Exploratory Data Analysis) to grasp the main characteristics of data, discover the relationships between features, and extract the important features affecting total order. For this, we will calculate the basic statistics of data with Python built-in functions, visualize data with functions from Matplotlib and Seaborn library, and implement more in-depth statistical calculation with SciPy library.

Once the analysis process is complete, we can feed the important features extracted from the analysis with a machine learning model to improve the model's prediction performance.